

## VOICE DATA TRANSMITTING AND RECEIVING SYSTEM

### BACKGROUND OF THE INVENTION

This application claims the benefit of Japanese Patent Application No. 2002-349621 filed on December 2, 2002, the contents of which are  
5 incorporated by reference herein.

The present invention relates to voice data transmitting and receiving systems and, more particularly, to a voice data transmitting and receiving system capable of securing meaning from data transmitted via a communication path, such as a quality of services (QoS) non-guaranteed  
10 network that may be, for instance, the Internet.

The Internet is in common use across borders and all over the world. Electronic commerce transactions and Internet telephone, i.e., internet protocol (IP) telephone, are attracting attention aside from such conventional applications as reading of web pages, electronic mail, and file  
15 transfer. This is greatly attributable to a rapid advancement of not only networks centered on line exchange in telephone networks, but also IP networks based on packet exchange.

In some IP telephone communication, various data including voice (or FAX) data and also data of still images and motion picture images, are  
20 converted to IP packets to be transferred in an IP based network. What is called Internet telephone is the utilization, in part of or full, of the same IP network, i.e., communication network, for communication in Internet protocol as is utilized for such applications as the World Wide Web, for voice telephone service utilizing IP network techniques.

25 Among IP telephone systems are the following three different systems. In a first one of these systems, voice messages are exchanged between personal computers which are dial-up interconnected on the Internet. In this system, it is necessary that the same software is installed

in the personal computers, which are in turn connected to a server. In a second system, communication cannot be obtained unless a telephone call is provided from a personal computer to a usual subscribed telephone set (converse call being impossible) or prearrangements are made

5 between the two sides. As a third system, two systems are present. In one of these systems, communication is made by inputting a user ID and a PIN via an Internet telephone gateway to a point of connection between an Internet network for communication between usual subscribers' telephone sets and a public telephone line switchboard. The other system is one for  
10 communication between direct Internet-coupled terminals. These systems are closest to the present telephone communication system, and their technical advancement is outstanding.

In the meantime, a system for transmitting a great deal of voice data in a narrow band has been proposed, in which on the transmission  
15 side an input voice is converted by voice recognition to character data, which are packeted and then transmitted, and on the reception side the received character data is converted to voice data, followed by voice synthesis and output of the resultant data as voice, thereby greatly reducing the transmitted data quantity and avoiding communication delay  
20 (see, for instance, Literature 1: Japanese patent laid-open Hei 10-285275). This system, however, although it has an advantage of reducing transmitted data quantity, is based on character data transfer. Therefore, the voice obtained by the synthesis has a fixed character, and is different in character from the speaker's voice.

25 In IP voice communication via an IP network such as the Internet or a local network without guaranteed QoS communication quality, usually Real Time Communication Packets (RTPs) of User Data Packet (UDP) protocol are used for transmission and reception of voice data. In this

case, although RTPs are used with importance attached to the real-time property of data in voice communication and motion picture playback, for the RTP no measure is provided against packet loss occurring on a communication path, and lost packets are not re-transferred, thus posing  
5 problems in voice quality, such as interruptions of voice.

To cope with these problems, heretofore, a system has been proposed, in which RTPs are transmitted together with preceding and succeeding packet data for an interpolating process according thereto, so that the voice will not be interrupted even in a packet loss event.  
10 However, in an environment in which data communication other than voice is frequently present, voice packet loss is pronounced, and the voice quality deterioration is too significant even by using the interpolation, sometimes resulting in failure of recognizing the meaning of the speech.

As shown above, the real-time voice communication by packet  
15 transmission is subject to loss of RTPs due to deterioration of a communication path environment, thus resulting in lost parts of voice data. Heretofore, satisfactory communication could be obtained only in good communication environments.

#### SUMMARY OF THE INVENTION

20 An object of the present invention, accordingly, is to provide a voice data transmitting and receiving system capable of allowing for the recognizing of the meaning of speech even in a deteriorated communication path environment.

Another object of the present invention is to provide a voice data  
25 transmitting and receiving system capable of allowing for the recognizing of the meaning of speech irrespective of packet loss due to causes in the communication path.

According to a first aspect of the present invention, there is

provided a voice data transmitting and receiving system for transmitting and receiving voice data as packet data via a network, wherein: on a transmission side voice clauses are divided and transmitted as packet data in divided clause units, and on a reception side the voice data is outputted  
5 as voice based on the received packet data in clause units.

According to a second aspect of the present invention, there is provided a voice data transmitting and receiving system, wherein: on the transmission side: real-time communication packets are generated based on input voice data; the input voice data is divided into clause units; and a  
10 plurality of RTP voice data in the clause units are transferred as packet data to a communication path; and on the reception side: packet data in clause units are obtained from packeted received data received via the communication path, thereby producing a replica of the RTPs in clause units; and outputting the voice data as voice based on the replica of the  
15 RTPs.

According to a third aspect of the present invention, there is provided a voice data transmitting and receiving system, wherein on the transmission side: real-time communication packets are generated based on input voice data; the input voice data is divided off into clause units; and  
20 a plurality of voice data RTPs in the clause units are combined into a single packet data and transferred to a communication path; and on the reception side: packet data in clause units are obtained from packeted received data received via the communication path, thereby producing a replica of the RTPs in clause units; and the voice data is outputted as  
25 voice based on the plurality of RTPs.

The data sent out from the transmission side is in the form of a file. On the transmission side either a re-transfer request is provided by recognizing the missing of received data or an interpolation process on the

received data is executed based on the received file data. The file data sent out from the transmission side is provided with discrimination data. In the reception, transmission side data is taken out from the received file data based on the discrimination data. The voice is divided into clauses  
5 based on voice recognition. The voice is divided into clauses based on an externally provided instruction. The voice is divided into clauses based on the sound level of the input voice. The voice is divided off into clauses based on changes in the input voice pitch level. The voice is divided off into clauses based on measured movement of the user's lips.  
10 The voice is divided off into clauses based on measured vibrations of the user's throat. The systems are selected based on the extent of communication per unit time between the transmission and reception sides.

According to a fourth aspect of the present invention, there is  
15 provided a voice data transmitting and receiving method as packet data via a network, wherein voice clauses are divided and transmitted as packet data in divided clause units in a transmission side, and the voice data is outputted as voice based on the received packet data in clause units in a receipt side.

20 According to a fifth aspect of the present invention, there is provided a voice data transmitting and receiving method, wherein: real-time communication packets are generated based on input voice data, the input voice data is divided into clause units and a plurality of RTP voice data in the clause units are transferred as packet data to a communication  
25 path in a transmission side; and packet data in clause units are obtained from packeted received data received for producing a replica of the RTPs in clause units; and the voice data is outputted as voice based on the replica of the RTPs in a receipt side.

According to a sixth aspect of the present invention, there is provided a voice data transmitting and receiving method, wherein real-time communication packets are generated based on input voice data, the input voice data is divided off into clause units and a plurality of voice data RTPs in the clause units are combined into a single packet data and transferred to a communication path in a transmission side; and packet data in clause units are obtained from packeted received data for producing a replica of the RTPs in clause units and the voice data is outputted as voice based on the plurality of RTPs.

Other objects and features will be clarified from the following description with reference to the attached drawings.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a system structure of a voice data transmitting and receiving system of a first embodiment according to the present invention;

Fig. 2 is a system structure of a voice data transmitting and receiving system of a second embodiment according to the present invention;

Fig. 3 is a system structure of a voice data transmitting and receiving system of a third embodiment according to the present invention;

Fig. 4 is a system structure of a voice data transmitting and receiving system of a fourth embodiment according to the present invention; and

Fig. 5 is a view for describing the operation of the embodiment shown in Fig. 4.

#### PREFERRED EMBODIMENTS OF THE INVENTION

Preferred embodiments of the present invention will now be described with reference to the drawings.

Fig. 1 is a system structure of a voice data transmitting and

receiving system of a first embodiment according to the present invention. In this embodiment, the transmission side comprises a communication terminal 11, a voice recognizer unit 12 and a packet combine unit 13, and the reception side which is connected to the transmission side via the  
5 Internet or like communication channel, comprises a packet division unit 21 and a communication terminal 22. While each user of course has both transmitting and receiving functions for the conversation purpose, in the following description the transmission and reception side are dealt with separately.

10 On the transmission side, a user's voice inputted to a microphone or like voice input device is processed as voice data in a communication terminal 11. On the reception side, a communication terminal 22 processes the voice, and outputs the processed voice via a loudspeaker or like voice output device.

15 On the transmission side, the communication terminal 11 generates real-time communication packets (hereinafter abbreviated as RTP) based on the input voice data. The voice recognizer unit 12 receives the voice data from the communication terminal 11, and executes a voice recognition process to divide off the voice into clause units. The  
20 packet combine unit 13 combines a plurality of voice data RTPs in clause units from the voice recognizer unit 12 into a single packet data to be sent out to a communication path. The packet combine unit 13 may send out the voice data RTPs in clause units as such.

On the reception side, the packet division unit 21 executes packet  
25 division of packeted received data received via the communication path to obtain RTPs of voice data in clause units, thus producing replica of a plurality of RTPs as clause units. The communication terminal 22 reproduces the transmission side voice data based on the plurality of

RTPs received from the packet division unit 21.

As shown above, in this embodiment clause units as divisions having means of voice composition are discriminated for transmission and reception as real-time communication packets in the discriminated clause units. Thus, even when packet loss occurs on the communication path due to deterioration of the communication environment due to such cause as communication line deterioration, the meaning of each clause can be transmitted, and reliable data transfer is possible.

A second embodiment of the voice data transmitting/receiving system according to the present invention will now be described with reference to the block diagram of Fig. 2. In Fig. 2, parts having functions like those in the case of Fig. 1 are designated by like reference numerals.

In this embodiment, the transmission side comprises a communication terminal 11, a voice recognizer unit 12, a packet combine unit 13 and a file producer (filing) unit 14, and the reception side which is connected to the transmission side via the Internet or like communication path comprises a packet division unit 21, a communication terminal 22 and a file receiver unit 23.

On the transmission side, the communication terminal 11 generates RTPs based on the input voice data. The voice recognizer unit 12 executes a voice recognition process on the voice data from the communication terminal 11 to divide off voice into clause units. The packet combine unit 13 combines a plurality of voice data RTPs in clause units to produce a single packet data to be sent out to the file producer unit 14. The file producer unit 14 produces as file of the receive packets, and sends out the file to the communication path.

On the reception side, the file receiver unit 23 receives the file data received via the communication path, and sends out the received file



data as packet data to the packet division unit 21. The file receiver unit 23 also recognizes missing, if any, of received data from the received file data in order to send out a data re-transfer request to the transmission side or execute an interpolating process on the received data so as to  
5 prevent missing of data.

The packet division unit 21 executes packet division of data received from the file receiver unit 23 to obtain the voice data RTPs in clause units and reproduce a replica of a plurality of RTPs as a single clause. The communication terminal 22 generates transmission side  
10 voice data from the plurality of RTPs received from the packet division unit 21, and causes the generated data to be outputted as voice from the loudspeaker.

In the above second embodiment, in addition to the advantage obtainable with the previous first embodiment that, transfer of the meaning  
15 of each clause and also reliable data transfer are obtainable even in the event of packet missing occurrence of the communication path due to deterioration of the communication environment stemming from such cause as communication line deterioration, it is possible to recognize missing of received data on the basis of the file data received from the file  
20 receiver unit 23 so as to send out a data re-transfer request or prevent missing of data through an interpolating process on the received data.

A third embodiment of the voice data transmitting/receiving system according to the present invention will now be described with reference to the block diagram of Fig. 3. In Fig. 3, parts having functions like those in  
25 the case of Fig. 2 are designated by like reference numerals.

This embodiment is basically the same in arrangement and operation with the above second embodiment shown in Fig. 2. This embodiment is greatly effective in a case where a firewall 24 is provided

between the communication path and the reception side. In this embodiment, the file producer unit 14 sends out the file data by making use of a generally open port such as HTTP and FTP, and in order for discrimination from any other file, discrimination data is provided after the  
5 file production.

On the reception side, the file receiver unit 23 which is connected via the Internet or like communication path, takes out a file transmitted from the file producer unit 14 from the full received file on the basis of the discrimination data, and sends out the taken-out file to the packet division  
10 unit 21. The file receiver unit 23, like the above case, recognizes missing of received data and sends out a data re-transfer instruction to prevent missing of data through an interpolation process of the received data.

In the third embodiment, in addition to the advantage obtainable in the first and second embodiments that transfer the meaning of each  
15 clause and also reliable data transfer is obtainable irrespective of packet missing on the communication path due to deterioration of the communication environment stemming from such cause as communication line deterioration, and that the data missing can be prevented by a data re-transfer process and a received data interpolation  
20 process based on the recognition of received data missing based on the received file data, it is possible to obtain communication with a communication terminal over a firewall.

Further embodiments of the present invention will now be described, which are different forms of a voice clause separation (or  
25 discrimination) system.

In a fourth embodiment of the present invention, signals representing manual clause divisions are outputted. It is thus possible to input necessary divisions with a person's judgment by using a manual

clause division device.

With this embodiment, division data can be inputted in any environment. Thus, the embodiment can be used not only for voice but also for music and continuous tones. Furthermore, the embodiment can  
5 be used for other RTP communications such as image communication.

In a fifth embodiment of the present invention, voice clause divisions are determined based on the measured input sound level. More specifically, the inputted sound level is measured, and an instant when the measured level is reduced down to a particular level, is determined to be a  
10 division or an off-division. The particular value in this case may be the noise level when the utterance comes to a pause.

This embodiment permits automatically dividing clauses at natural divisions in the utterance.

In a sixth embodiment of the present invention, voice clause  
15 divisions are determined based on the measured input sound pitch. Specifically, the input sound pitch is measured, and an instant when a pitch difference exceeds a constant value is determined to be a division.

This embodiment permits automatically discriminating off-divisions of utterance irrespective of whether the background noise level is high.

20 In a seventh embodiment of the present invention, voice clause off-divisions are determined based on the movement of lips by making image measurement of the face of a person during voice input. In other words, image measurement of a person during voice input is made, and an instant when the movement of the lips becomes stagnant is determined  
25 to be a division.

With this embodiment, divisions are determined with a mechanism different from the voice process, and it is thus possible to automatically discriminate divisions without any appropriate voice discriminator.

In an eighth embodiment of the present invention, voice clause divisions are determined based on the measured vibrations of the throat. Specifically, vibrations of the throat are measured, and an instant when the vibration is stopped is determined to be the division.

5           With this embodiment, divisions-off are determined with a mechanism different from the voice process, and it is thus possible to automatically discriminate off-divisions without any appropriate voice discrimination. The embodiment can also be used in the case of extremely low voice level.

10           In a ninth embodiment of the present invention, voice clause off-divisions are determined by a method of discrimination and analysis of voice as compositions. Specifically, voice is analyzed as compositions, and proper off-divisions are determined. As techniques of analysis of voice to compositions, well-known techniques may be used.

15           With this embodiment, it is possible to automatically determine off-divisions from meanings even in an environment, in which the above method cannot be utilized, for instance with flat and long continuous voice.

A tenth embodiment of the present invention will now be described.

20           This embodiment is installed on the transmission side (or reception side), and an optimum communication means is informed to both the transmitting and receiving communication terminals by observing the communication status.

Fig. 4 is a block diagram showing the present embodiment.

25           This embodiment comprises a transmission/reception monitor unit 31 for sensing the start and end of communication, a communication time storage unit 32 for accumulating communication time, a communication extent storage unit 33 for accumulating the quantity of transmitted or

received data, a reference value/corresponding means storage unit 34 for  
storing reference values for switching communication means and also  
these communication means, a comparative computer unit 35 for  
calculating the communication extent from the outputs of the  
5 communication time storage unit 32 and the communication extent storage  
unit 33, comparing the calculated value with the reference values stored in  
the reference value/corresponding means storage unit 34, and a  
communication means informing unit 36 for receiving a communication  
means from the comparative computer means 35 and commanding the  
10 switching to the received communication means.

The operation of the embodiment will now be described with  
reference to Fig. 5.

When the communication is started, the transmission/reception  
monitoring unit 31 senses the start for communication, and causes the  
15 communication time storage unit 32 and the communication extent storage  
unit 33 to start accumulations, respectively. Whenever a constant time  
passes, the data stored in the communication time storage unit 32 and the  
communication extent storage unit 33 are sent out to the comparative  
computer unit 35, while the accumulated data in the communication time  
20 storage unit 32 and the communication extent storage unit 33 are deleted.  
The comparative computer unit 35 computes the extent of communication  
per unit time from the data sent out from the communication time storage  
unit 32 and the communication extent storage unit 33, compares the result  
of calculation with the reference values stored in the reference  
25 value/corresponding means storage unit 34, and sends out the data of the  
corresponding communication means to the communication means  
informing unit 36. The communication means informing unit 36 sends out  
a command for switching to the selected communication means to the

communication terminal. When the communication is ended, the transmission/reception monitoring unit 31 senses the end of accumulations, and notifies the end of accumulations and deletion of the stored values to the communication time storage unit 32 and the communication extent  
5 storage unit 33. As shown, in this embodiment the above systems can be selectively used based on the extent of communication per unit time between the transmission and reception sides.

With this embodiment, the transmission/reception can do communication with optimum communication means matched to the  
10 environment of the communication path. As examples of the communication means, the RTP communication may be normally selected, the clause division packet communication may be selected in a bad communication path environment, and the file production communication may be selected in the worse communication path environment.

15 The arrangements and operations of the preferred embodiments have been described above. However, these embodiments are merely examples of the present invention and are by no means limitative. It will now be readily understood to the person skilled in the art that various changes and modifications are possible in dependence on particular uses  
20 without departing from the scope of the present invention.

As has been described in the foregoing, with the voice data transmission/reception system according to the present invention not only is it possible to obtain transfer of the meaning of each clause and reliable data transfer even when missing of packets occurs on the communication  
25 path due to deterioration of communication environments stemming from such cause as communication line deterioration, but also it is possible to send out a data re-transfer request by recognizing missing of received data and prevent data missing by a received data interpolation process,

thus permitting communication with a communication terminal in the presence of a firewall.

Furthermore, the transmitting and receiving communication terminals can do communication with proper communication means by  
5 matching the environment of the communication path.

Changes in construction will occur to those skilled in the art and various apparently different modifications and embodiments may be made without departing from the scope of the present invention. The matter set forth in the foregoing description and accompanying drawings is offered by  
10 way of illustration only. It is therefore intended that the foregoing description be regarded as illustrative rather than limiting.